And How Experiments Begin: The International Prototype Kilogram and the Planck Constant

Sally Riordan

University of Cambridge

sr206@cam.ac.uk

Department of History and Philosophy of Science
University of Cambridge
Free School Lane
Cambridge
CB2 3RH
United Kingdom

And How Experiments Begin: The International Prototype Kilogram and the Planck Constant

Abstract

The artefact that has defined the kilogram since 1889 is to be retired and the kilogram will instead be defined by fixing the value of the Planck constant. In this paper, I detail some of the elements of this reform, believing that the case study should prompt philosophers to reassess the role scientific standards play in the progress of the physical sciences. A metrological account of scientific standards should explain metrology's more theoretical motivations and also acknowledge its empirical contribution to the physical sciences. I present three theses towards this end. I develop a more thoroughgoing and yet much weaker version of Bridgman's operational attitude. I present a picture of the physical sciences united by metrology. Finally, I present the case for a quiet form of realism that attempts to accommodate both the more theoretical and the more pragmatic motivations of the metrologist.

Keywords

conventionalism, kilogram, measurement, metrology, operationalism, SI

1. The conventionalist account of scientific standards

A straightforward view of scientific standards goes as follows. Equipped with an understanding of what is being measured (length), it becomes necessary, if experiments are to be conducted at all, to "point" to a certain quantity of that substance (one metre). The pointer, or definition of the standard, is given by a stipulation (the distance between the two lines upon this metal bar, when at the temperature of melting ice, shall be one metre). At least one procedure, whether tacitly understood or developed specifically for the occasion, must be available in order to then realise this pointer (the lining a metal bar of approximately one metre against the prototype metre and taking a reading of its length from a measuring microscope). A particular realisation may be improved to increase its accuracy (giving a value closer to that determined by the pointer), to increase its precision (giving closer values when the procedure is repeated within a short period of time), to increase its stability (giving closer values when the procedure is repeated after a long period of time), or to increase its reliability (giving closer values when the procedure is repeated in a different laboratories, by less skilled operators, in slightly different ways or using different equipment). The most plausible reason for discarding a pointer altogether, on this straight-forward view, is that an alternative stipulation promises realisations of increased accuracy, precision, stability or reliability. Less drastically, a pointer itself may be honed in order to improve its realisations ("this bar being subject to standard atmospheric pressure and supported on two cylinders of at least one centimetre diameter, symmetrically placed in the same horizontal plane at a distance of 571 mm from each other" (CGPM, 1928, 49)). All of this is the work of the metrologist, who responds to demands from the rest of science (and beyond) for increased accuracy, precision, stability and reliability in scientific standards by improving existing pointers and their realisations, as well as developing new ones.

This straightforward view of scientific standards is one that makes sharp distinctions between what is being measured, the definition of a unit of that whatness, and the method (or methods) by which it is realised. The distinctions are most easily played out by supposing that science divides neatly into three activities: theoreticians generate models of nature from the evidence before them; metrologists develop the measures required to test those models; experimentalists conduct the tests; theoreticians respond to the results with new models… The metrological stage is, however, so often overlooked that of these three transitions, only that between the theoretician and experimentalist has been analysed in depth by philosophers.

On the straight-forward view of scientific standards, it is, in addition, taken for granted that metrology responds to (but does not directly inform) scientific theory: our understanding of our measures comes after and makes no contribution to, our understanding of what we are measuring. A conceptual change driven by scientific theory can thus only be acknowledged in metrology by replacing a pointer ("Because we have a better understanding of what *length* is, we will now define a metre by the speed of light"). It is also the case that metrology provides for, but does not directly respond to, the work of experimentalists. On the straight-forward view, an experiment can only begin once the metrological work required for that experiment has come to an end. The instruments used by the experimentalists are created and calibrated elsewhere. Although it has been appreciated generally that scientific activity, progress and knowledge cannot be neatly divided into the theoretical and experimental, the consequences for how we view the place of metrology in science have not perhaps been so thoroughly considered. The assumption of the isolation of metrology (that is does not contribute to scientific progress in profound ways) results from juxtaposing realism regarding the posits of science with a conventionalist about the standards of science; it is one that can occasionally can be seen operating within the scientific community. For example, it is perhaps one of the reasons why it was difficult to recruit graduate scientists to metrological work in the 1960s (Quinn, 2011).

It is at least true that, outwardly, metrology does not appear to be a theoretical science in the same way as the physical sciences it supports; lacking its own models of nature, metrology stands apart. Even a cursory glance at the history of metrology furnishes us with evidence for the lag between theoretical world of science and its metrological afterthought: after the onslaught upon our concept of mass by twentieth-century physics, for example, the scientific community remained tied to realising mass by the pointer that had been manufactured before Albert Einstein's birth and officially accepted in the early years of his childhood. On the straight-forward view, how we define and measure a unit doesn't matter to our understanding of the quality that unit represents. The decisions regarding the scales against we measure it—and thus the choice of pointers and methods of realisation—are entirely matters of convention, for it is a choice guided by human values and not nature. The sharp distinction between the theoretical study of a quality and the conventional methods of measurement places metrology in its entirety in the "subjective", "arbitrary" or "conventional" bin. I call it *the conventionalist account of scientific standards*.

Despite the name, this viewpoint does not obviously sit astride a broader—a more generic—conventionalism about scientific concepts, of the kind associated with the likes of Ernst Mach, Henri Poincaré, Percy Bridgman, Rudolf Carnap or Hans Reichenbach. The conventionalist view of scientific standards does not capture the thinking of these more careful conventionalist thinkers and, as we shall see, is ironically in deep conflict with generic scientific conventionalism. It is, in fact, more neatly accommodated by a naively realist view of the physical sciences: a quantity sits in nature, regardless of whether and how we decide to measure it. (I appreciate that the label "conventionalist" is therefore confusing. However, I hope that my terminology will not be unfortunate in the long run, because it will encourage us to pick apart different threads of conventionalist thinking about science, not necessarily mutually supporting.) When lined up with naïve realism, as well as the assumption that the theoretical and the experimental can be cleanly distinguished in science, it no doubt appears that I have prepared the perfect straw-man for myself. However, I do not hold the view that the conventionalist account of scientific standards is always wildly inaccurate or always inapplicable. I do not deny its usefulness and—if pressed to talk about such things—I would admit that it contains an element of truth. I accept the account as one model, albeit an elementary and particularly unenlightening one, regarding the connection between scientific standards and physical theory. And I will be pointing out the advantages it has, when dealing with scientific standards, over more generic conventionalist accounts. The theses I present here may seem radical alternatives, but I believe that no account of scientific standards can ever be more than an iteration of the conventionalist's. My suggestion is not that these distinctions are to be done away with altogether, but that they should be seen for what they are, conventions themselves. My first intention in articulating the conventionalist account of scientific standards, then, is to highlight its prevalence in philosophical and more general thinking and our unwitting reliance upon it. It is easy—perhaps too easy—to interpret metrological history through its lens. As a result of its hold on our thinking, it leads both realists and anti-realists astray: having appreciated the theory-ladenness of experiment, there remains no interesting question to be asked regarding the role of scientific standards in physical theory. Despite, then, what I have already said regarding the resonance between the conventionalist account of scientific standards and naïve realism, I argue that it is also the entrenched view (albeit somewhat warped) within anti-realist philosophies. For example, it is because anti-realists did not more comprehensively reject the conventionalist account of scientific standards that the philosophical community was burdened with a doomed form of operationalism, limping before the race to explain the meaning of scientific terms began. Whether we have noticed it or not, the conventionalist account of scientific standards is with us, in all its damned simplicity. In § 2, I present an overview of modern metrology and the current reform to the international system of scientific units (SI), demonstrating how natural it is, even for metrologists themselves, to assume the position of a conventionalist with regard to scientific standards.

My second intention is to pinpoint where the conventionalist account of scientific standards fails to account for some of the finer details of metrological practice and to therefore indicate what a *metrological account of scientific standards* might look like. I use the term generically to refer to any account that pays attention to such details. The BIPM define a 'measurement standard' as a, "realization of the definition of a given quantity, with stated quantity value and associated measurement uncertainty, used as a reference" (2012, 46). I use 'scientific standard' because, as will become clear (§ 5), I view standards as the repeatable procedures, ratified by metrologists, that produce these realisations. If this thesis has a constructivist shimmer to it, it is complemented by the thesis that metrological ratification is tightly constrained by experimental realities (§ 3). I do not, however, defend a particular position here, but suggest three theses that a metrological account of scientific standards might incorporate. The case study at hand is the current redefinition of the kilogram. I describe here some of the details of this reform: the motivations that led to the change (§ 3); the development of the Watt balance (§ 4); and the procedures for cleaning and washing kilogram prototypes (§ 5). The underlying theme to the analysis is to consider how the clean-line thinking of the scientific-standard conventionalist falls apart in the face of these details: between precision and accuracy (§ 3); between calibration and experiment (§ 4); and between meaning and method (§ 5). In each of these three cases, I explore an extreme response to the case study, before settling upon a weaker statement of it, in an attempt to develop a metrological account of scientific standards.

In § 3, I argue that the desire for stability in scientific standards has a theoretical aspect that cannot merely be explained as the requirement to reduce the uncertainty of a standard in the long-term. The motivation for the redefinition of the kilogram is both pragmatic and theoretical in nature. One conclusion available to the analyst is to reject the naïve realism associated with the conventionalist account of scientific standards, instead recognising a milder and yet more specific realism. Roughly, *metrological realism* is the thesis that uncertainty of measurement does not just indicate the practical and technological limitations of empirical inquiry, but also measures the limits placed upon science by the way the world is. In attempting to reduce the uncertainty of measurement, then, metrologists are investigating and modelling nature.

In § 4, I explain how the Watt balance, once used to determine the Planck constant, will instead be used to realise the kilogram at the highest level. At a minimum, this indicates that the choice of metrology determines the lines between experiment and calibration. A stronger conclusion available to the analyst is to suppose that an experiment includes all the chains of calibration that are associated with its determinations. Giving calibration the same epistemic status as experimentation, and perhaps metrology the same as other physical sciences, I propose the *metrological unification of the physical sciences*. Roughly, this is the view of the physical sciences as the collection of all physical determinations, whether of experiment or of calibration, together with the theoretical associations of these determinations.

In § 5, I describe the current official procedure for cleaning and washing kilogram prototypes. The study indicates that there is no distinction between a pointer (the definition of a standard) and its *mise en pratique* (its official method of realisation) that is of epistemic importance. This paves the way for considering that a *mise en pratique* contributes to the meaning of a scientific standard. To put the point a little too provocatively: the way we measure the world contributes to the meaning of science. The point leads us to consider operationalism and again raises the problem of accommodating metrological progress from a conventionalist point of view. In response to this, I develop a more thoroughgoing, but yet much weaker version, of Bridgman's operational attitude. Roughly, *metrological operationalism* is the thesis that measurement procedures contribute, at least partially, significance to scientific standards and the quantities associated with these standards.

Altogether, this points towards an (as yet hazy) metrological account of scientific standards. Metrology is more theoretical, empirical and meaningful than the conventionalist account of scientific standards allows. I speculate about the connections between these theses in § 6. Regardless of the details of my suggestions, however, I hope that the question regarding the role of scientific standards in physical theory now presents itself as an interesting one. To begin the analysis, I must describe the deep hold the conventionalist account of scientific standards has on the story of metrology.

2. A conventionalist perspective upon the history of modern metrology

There are good reasons to take the modern age of metrology to begin in 1875, when representatives from seventeen nations met in Paris to sign a treaty. The Metre Convention created the political structure through which international agreement on matters of scientific standards has since been made: it brought about the International Bureau of Weights and Measures (BIPM), an international research organisation working to improve scientific standards, and its governing body, the General Conference on Weights and Measures (CGPM), a quadrennial meeting of delegates from member governments which directs the research at the BIPM, as well as the International Committee for Weights and Measures (CIPM), eighteen individuals from member states, who make recommendations to the CGPM. The metric system itself dates back to revolutionary France of the 1790s, when the metre was defined by a portion of the Earth's circumference and the kilogram was defined by a volume of water at its densest. With these definitions, however, the original creators of the metric system had intended to establish natural scientific standards, beyond the reach of human error and without the need of human maintenance.[i] Although the original creators expected that it might be one day necessary to replace the platinum-iridium artefacts that exemplified the new standards, they believed that the real work of length and mass metrology had been completed. The nineteenth century witnessed the growing recognition that the metric project had failed in this intention; the understanding that scientific standards develop alongside the rest of science brought about the founding of the BIPM and metrology as we know it today.

As the name suggests, the original purpose of the Metre Convention was to advance length metrology. The International Metre Commission, which led directly to the Metre Convention, was assembled in 1869 in response to the difficulties faced by the international geodesic community in establishing agreement and uniformity in length measurement.[ii] In addition, it was appreciated that a more precise calibration of length could be obtained against a standard defined by two fine lines marked upon a metal bar, instead of one defined by the entire length of a bar. The French section of the commission initially understood their remit to be limited to the creation and distribution of replicas of the original metre prototype, both end- and line-standards. The project was soon shaped, however, by the international view that the original artefact was to be replaced. And because there was a metric connection between mass and length—the kilogram had originally been defined by a *decimetre* cube of water at 4 ºC and the original kilogram prototype had been made at the same time as that of the metre—the commission nervously took responsibility for creating a new kilogram artefact at the same time. Doubts were expressed that they had the power to do so, yet members of the commission voted, by the narrow margin of ten votes to eight, to take on this additional project. The practical demands of international geodesy brought about the creation of new prototypes of both the metre and the kilogram.

Modern metrology was thus born in a whirlwind of conventionalism, acknowledging that scientific standards were necessarily designed, made and looked after by metrologists according to the demands of the scientific community. And the conventionalist viewpoint was not restricted to the metrological community, but was reflected and a reflection of the conventionalist views of physicists and philosophers at this time. In the period 1875–1930, metrology held a high status within both physics and philosophy of physics. Metrological work attracted the attention of successful physicists in the U.K. (including James Clark Maxwell, Lord Kelvin, J. J. Thomson) and was rewarded with Nobel prizes (Alfred Michelson in 1907 and Charles Guillaume, director of the BIPM, in 1920).[iii] The philosophy of the period was informed by metrological practice (before his rise as a philosopher, for example, C. S. Pierce had contributed to measurements of the intensity of the earth's gravitational field). The rising philosophies of science, whether labelled pragmatic, logical, empirical or conventional, paid attention to the conventional nature of metrology (most notably, I'll be turning to the work of Reichenbach in § 4).

As the twentieth century progressed, one of the more salient features of metrological progress was an increasing rigour, arising not least as a result of the work of the BIPM and reflected in increasing standardisation, efficiency and formality. Developments of this kind are readily interpreted from the conventionalist's point of view. In my first example of this interpretation, consider that, during the twentieth century, it became standard metrological practice to associate each measurement in the physical sciences with a parameter that "characterizes the dispersion of the values that could reasonably be attributed to the measurand" (BIPM, 2008, 2). The estimation of uncertainty is not entirely algorithmic. It does include a statistical analysis of the experimental data and therefore takes into account the variation of data points. (Thus uncertainty is, strictly speaking, neither attached to a measurement procedure in general, nor to a single enactment of it; uncertainty is, in the first place, associated with a limited number of repetitions of a procedure, which together constitute one measurement.) Uncertainties calculated statistically are called "Type A" uncertainties; in addition, estimates are made of non-statistical, "Type B" uncertainties, which result from other sources, including calibration certificates, manufacturer's specifications, and from common sense (Bell, 2001, 11). A *combined standard uncertainty* is calculated from the (potentially many) Type A and Type B uncertainties (BIPM, 2008, 7). The conventionalist account of scientific standards encourages us to view metrological progress, including the current reform to the SI, as an attempt to decrease the uncertainty with which we can make measurements and thus to view the combined standard uncertainty as a measure of the imprecision, instability, unreliability and inaccuracy of measurement. Most obviously, it is a measure that, in its inclusion of Type A uncertainties, relates directly to the imprecision with which a standard has been realised. In theory, statistical analysis of data taken from different laboratories and across time periods can also reflect the unreliability and instability of a standard's realisation. In the case of the IPK, metrologists have attempted to quantify the uncertainty associated with its instability; the conventionalist account of scientific standards appears to be supported by the metrologists who reason that the last artefact from the SI must be given up because of its "inherent instability" (Davies, 2005, 2263). Further indicators of unreliability are represented by Type B uncertainties, including, for example, the calibration of instruments used in the measurement and dependencies upon external experimental values. The demand for an accurate standard, on the other hand, is dealt with in a less direct way by combined standard uncertainty. It is necessary to infer the accuracy of a measurement from the coherence of results obtained using methods that are theoretically different; the combined standard uncertainty merely provides a range of values from which it can be determined whether results are compliant.

A second example of interpreting the twentieth century metrology through the conventionalist lens comes from considering the increasing formality of metrological practice during this time. The BIPM came to recognise its responsibilities for determining and communicating each *mise en pratique*, a procedure by which an SI standard is realised "at the highest level" (CIPM, 2008, 62). The term applies to the procedures for realising a standard using an artefact (regarding the handling of the IPK, for example) and to those for realising a standard from a fundamental constant (regarding the operation of the Watt balance, for example). The BIPM now takes great care when choosing the wording of a *mise en pratique*, as well as that of the definition of a standard, and distinguishes between the two. It recognises, in fact, that the distinction has been made especially sharp by the current reform of the SI, in which the kilogram, the ampere, the kelvin and the mole are being redefined by fixing the values of four physical constants (the Planck constant, the elementary charge, the Boltzmann constant and the Avogadro constant, respectively), in addition to the three physical constants whose numerical values have been fixed to date (the caesium hyperfine frequency, the speed of light in vacuum and the luminous efficacy of a defined radiation). Thus, at the core of the reformed SI, are fixed values for seven physical constants of nature, which correspond to seven standards of measurement (given in Table 1). In marked contrast to these seven numbers, are the methods being developed that will stand as *mise en pratiques* for the seven associated units. Furthermore, the BIPM (2013, 9–10) makes the distinction precisely for the reason of reducing uncertainty in the long term:

> "The use of seven defining constants is the simplest and most fundamental way to define the SI […]. In this way no distinction is made between base units and derived units; all units are simply described as SI units. This also effectively decouples the definition and practical realization of the units. While the definitions may remain unchanged over a long period of time, the practical realizations can be established by many different experiments, including totally new experiments not yet devised. This allows for more rigorous intercomparisons of the practical realizations and a lower uncertainty, as the technologies evolve."

Beneath some statements of the BIPM, regarding its intentions and its purpose, then, it is possible to interpret a conventionalist view of scientific statements. Metrologists redefine pointers, as well as the realisations for those pointers, in order to increase the precision, reliability, stability (and perhaps accuracy) of scientific standards. The single aim of metrology is to reduce the uncertainty with which measurements can be made. On a first pass, modern metrology—including the current reform of the SI—conforms to the conventionalist account of scientific standards.

3. Motivations for a new definition of the kilogram and the case for metrological realism

In 1889, the original exemplar of the kilogram was replaced by a platinum-iridium artefact manufactured in London by George Matthey. The International Prototype Kilogram (IPK) was accepted as the kilogram standard at the first meeting of the CGPM (1890, 34): "This prototype shall henceforth be considered as the unit of mass." More than anything else, what marked the new kilogram standard from the Kilogram of the Archives was that it was one of 42 prototypes, all made to the same specifications. In fact, the creation of the new metric standards under the direction of the newly-established BIPM had been delayed during the period 1774–1882 because of the desire to make all the prototypes from a single casting of platinum-iridium, a requirement that was eventually dropped (Quinn, 2011, Ch. 4). The commissioners believed it of importance that the prototypes be as similar as possible because they were aware that, in creating many prototypes of the new standards, they were creating a check upon the stability of the IPK. It was envisaged that the 42 prototypes would be brought together at intervals to recalibrate them against the IPK. The first comparisons were performed before the prototypes were distributed to governments worldwide, in the period 1886–1889, by Max Thiesen and Damian Kreichgauer. Each prototype was compared directly with the IPK, as well as with twelve other prototypes. From a statistical analysis of the variations in the measurements taken, it was concluded that the mass of each prototype with respect to the IPK was accurate to within 0.002 mg (Thiesen, 1898, C17). As we have seen, the concept of uncertainty has since developed. This analysis would thus not be accepted today: consideration of additional causes of uncertainty would result in a higher measure (Quinn, 2011, 123).

The second periodic verification of the national prototypes was conducted by Albert Bonhoure, in the period 1946–1953 (CIPM, 1946, 171–178). This included additional prototypes, numbered between 44 and 55 (non-inclusively), made by Matthey's firm as the membership of the Metre Convention increased in the first half of the nineteenth century. The third periodic verification of the national prototypes was conducted by Georges Girard, in the period 1988–1992 (CIPM, 1993, G35–G50). Again, the collection of prototypes had expanded; the new additions were numbered between 56 and 65 (non-inclusively). Altogether, the third verification measured the mass of 51 kilogram prototypes with respect to the IPK (listed in Table 2): the six official copies of the IPK (K1, No. 1, No. 7, No. 8(41), No. 32, No. 37 and No. 38), the three working copies of the IPK belonging to the BIPM (No. 9, No. 25, and No. 31), and forty-two national prototypes (most of which were held by governmental metrological laboratories).[iv]

In all, the third verification confirmed what had already been showing in the results of the second: in 1992, the mass of an original copy of the IPK was, on average, 25 μg more than it was in 1889 (shown in Figure 1); the mass of a secondary copy of the IPK was, on average, 40 μg more than it was in 1946 (shown in Figure 2). One way of interpreting this data is to suppose that the mass of the IPK is itself drifting, losing mass at a rate of the order of 0.5 μg per annum. The metrological community itself, however, has shown some caution in reaching this conclusion. Metrologists appreciate that their current knowledge of the behaviour of platinum-iridium prototypes is too limited to determine the causes of the apparent drift (Quinn, 2011). It is plausible, for example, that the discrepancies between the

IPK and the national prototype are due to the response of the IPK to washing and cleaning: Girard's tests demonstrated that the IPK changed in mass more markedly after washing and cleaning than other prototypes (shown in Figure 4).[v] What is clear is that, over long periods of time, the prototypes are unstable to an extent that can be detected.

It was not, however, this result that prompted the current reform of the mass standard. The apparent drift of 0.5 μg per annum in the kilogram standard has not yet created any practical problems for commercial, industrial or scientific users of the kilogram. (To put the drift into perspective, consider that Girard (1990) judged that the mass of each national prototype had been measured against the IPK with un uncertainty of 2.3 μg; such mass calibrations are the most certain that can be done in SI units because all other mass measurements must take into account this uncertainty.) Furthermore, it has long been presumed by metrologists that the mass of a metal piece changes over time; the result of the third verification of the IPK held little surprise. If the drift had not been apparent, the desire to replace the IPK would remain. Indeed, the desire to replace an artefact mass standard with something "more fundamental" existed long before the third—and even the second—verification took place. The third periodic verification of the national prototypes was conducted, in fact, because the metrological community could not achieve what it really wished to. The fifteenth CGPM of 1975 requested that the BIPM conduct the verification, as well as continue research into the improvement of mass standards comparisons, for the reason, "that there is no immediate prospect of defining the mass unit in terms of atomic constants with a comparable precision" (1975, 103–104).

The desire to define a mass unit in more fundamental terms than by a metal artefact is older than the technological ability and the practical requirements to do so. It is a sentiment that echoes throughout the metrological literature: in the papers and review articles of *Metrologia* since the journal's inception in 1965; in the annals of the CGPM, produced every four years since 1889 and the associated minutes of the CIPM's yearly meetings since 1876; in the parliamentary archives of the French Revolution concerning the creation of the original metric system; and in the works of natural philosophers of the seventeenth and eighteenth centuries.[vi] The Metric Convention turned out to be only a temporary postponement of this goal. The concept of a fundamental scientific standard is one that has never gone away entirely, although it has changed in this time, most obviously in response to a changing body of scientific knowledge. A fundamental standard has usually been understood to be in some way given or representative of nature (an idea that has been associated with immutability, uniformity, self-maintenance, reproducibility, deliverance by experiment and explainability to aliens).[vii] Since the mid-nineteenth century, however, it has increasingly been understood to be a molecular or atomic standard, one that relies upon counting phenomena in the microscopic realm. One of the earliest and clearest articulations of this point was made by James Clerk Maxwell (1870, 7):

> "If, then, we wish to obtain standards of length, time, and mass which shall be absolutely permanent, we must seek them not in the dimensions, or the motion, or the mass of our planet, but in the wave-length, the period of vibration, and the absolute mass of these imperishable and unalterable and perfectly similar molecules."

By the turn of the twentieth century, delivering fundamentality via atomic standards had become a possibility. This was recognised by the astronomer David Gill in his presidential address to the British Association for the Advancement of Science in 1907. Gill acknowledged that he had been influenced by Maxwell's dream of a standard communicable to aliens during an 1859 lecture Maxwell gave in Aberdeen. He also recognised the very practical issue of breaking or losing an artefact standard. It was scientific discovery that had seemingly revealed the answer to these issues, nature herself directing metrologists toward an atomic metrology. Thus, Gill (1907, 195) explained that the International Prototype Metre was not scientifically described by the label "one metre" but only,

> "as a piece of metal whose length at 0º C. at the epoch A.D. 1906 is 1,553,164 times the wave-length of the red line of the spectrum of cadmium when the latter is observed in dry air at the temperature of 150 ºC. of the normal hydrogen-scale at a pressure of 760 mm. of mercury at 0º C."

The description made use of the most recent determinations of the red-light wavelength emitted from cadmium, by Alfred Pérot and Charles Fabry in 1906.[viii] From Gill's point of view, there was a better (a more scientific) description of "one metre" than that given by the International Prototype Metre. It has not solely been a desire to decrease the uncertainty with which artefacts can be compared that has driven SI reform, for the challenge comes from the opposite direction. Since the creation of the IPK, metrologists have wanted instead to decrease the uncertainties with which a mass standard could be realised in alternative, more fundamental ways. The prospect of doing so eventually came from two independent lines of development. On the one hand, Bryan Kibble developed a balance that compared gravitational with electrical power; developments in quantum electrodynamics then enabled a theoretical link to be made between a gravitational mass and quantum constants. On the other hand, computing technologies brought the possibility of measuring mass by counting silicon atoms. In both cases, metrologists at the turn of the twenty-first century were working to improve the uncertainties of these techniques, aiming only to match the uncertainty associated with calibrating against the IPK, and not to improve upon it. The target uncertainty was accepted to be 2 in 10^8 parts, a measurement of a kilogram to within 20 μg (Kelley, 2001, 860).

The reform to the kilogram was not, then, immediately driven by a desire to decrease the uncertainty with which platinum-iridium prototypes can be compared. The most obvious first responses to the apparent drift of the IPK with respect to its copies, after all, is to investigate the causes of the drift (a line of research that has been left relatively unexplored by the metrological community) or to define the kilogram by the average mass of a collection of pieces (a possibility that has not seriously been considered since the installation of the IPK).[ix] The progress of metrology is occasionally driven by more theoretical concerns, which can be interpreted as an expression of a scientific realism of sorts. The apparent drift between the IPK and its prototypes is, after all, only a relatively minor part of a larger problem regarding the long-term stability of platinum-iridium artefacts. There remains the further possibility of a more unsettling drift, in which all the platinum-iridium pieces are drifting from their original masses. The metrological community accepts that the masses of the prototypes changes in the long term and estimates this change (without experimental evidence) to be within ten times that of the drift between the pieces themselves, resulting in an upper limit of 5 μg per year (Davis, 2008; Quinn, 2011). Such estimates assume that the mass of the iridio-platinum prototypes is to be compared to a more accurate indicator of mass. There is, however, no experimental warrant for believing that the Planck constant will provide more stability to mass measurement in the longer term. The turn towards the fundamental constants is founded on Gill’s assumption, seemingly both innocuous and true: the natural world (and not just the practical requirements and realities of experimental science) determines better and worse ways of defining scientific standards.

Conventionalism regarding scientific standards is naturally coupled with a naïve realism regarding the posits of scientific theory and can, in this way, account for the theoretical motivations of metrological progress. It does this, however, with a heavy hand. There is the temptation to go further than is strictly warranted by the details of metrological practice and assume the existence of fundamental standards in nature. In comparison, the realism called for by Gill’s assumption makes only mild ontological commitments.

Furthermore, because the conventional work of the metrologist is crudely aligned with the investigative work of the theoretician, it is assumed that the kilogram is only reformed once a new understanding of mass is delivered as a result of progress in the physical sciences. The suggestion to fix the value of the Planck constant was, however, one that emerged within metrology. The thesis that the Planck constant is invariable across space and time is not one that is thoroughly integrated into theoretical physics: the Copernican principle that the laws of physics are the same for all observers is compatible with smooth changes in the fundamental constants and physicists continue to speculate whether the fundamental constants change over time. The point is most famously expressed by Dirac (1937), but remains a part of contemporary physics (Avetissian, 2009). It remained for metrologists to test the stability of the Planck constant and perform the precision measurements required to demonstrate it provided a suitable grounding for a scientific standard (Steiner *et al*, 2005; Eichenberger *et al.*, 2009). It was not because science had progressed to the stage that the uncertainty of the Planck constant was on a par with mass measurements that the SI reform was initiated, but the reverse: it was metrologists who asked how well the constant is known and determined to make further precision measurements of it. My claim here is not just that the line between calibration and experiment blurs, but so too between metrology and precision measurement, often considered a part of the physical sciences proper. The experimental and theoretical work of determining the constancy of the Planck constant is not done separately from metrology.

Furthermore, an alternative proposal for defining the kilogram, involving the counting of silicon atoms, was rejected, despite it being recognised as closer to our intuitive understanding of what mass is and thus “more readily comprehensible” (Hill *et al.*, 2011). We shall see that the Watt balance weighs masses using electromagnetic units, but there is no drive from theoretical physics that encourages metrologists to suppose that the true essence of mass is explained by electromagnetic theory. It was for the metrological community to determine how to interpret mass for the purposes of measurement. Ironically, it relies upon the quantification of precision and stability in order to do this: these measures are not merely indicators of the practical difficulties faced by the users of the kilogram, but indicators of what nature presents as a better measure of reality.

The metrological investigation of the invariance of the Planck constant and of material artefacts is only just beginning. As part of the current reform, metrologists are currently creating a collection of kilogram artefacts to replace the national prototypes. It was important, when making the last collection, that they were as similar as possible. It is important, in the current work, to include kilograms of many different materials. The purpose will be to monitor the stability of artefacts against the Planck constant. It is plausible that there are surprises are in store for us regarding the stability of mass measurements and, as a result, our understanding of how best to represent mass. But in any case, the responsibility for this lies firmly in the hands of the metrological community. That responsibility includes determining the status of the statement, “The Planck constant does not change in space or time,” and its role in physical theory. Metrologists have the power to underline it as an important theoretical principle, or leave it as an empirical possibility, but it remains open to the results of precision testing. It won’t do, then, to consider the metrological choice as entirely or merely conventional. In comparison to the realism that is usually coupled with scientific-standard conventionalism, the realism called for by Gill’s assumption makes strong epistemological demands.

The motivation for the stability of a standard is thus a ‘thick’ one: as well as being a practical desire that our standards do not change over time, it is an indication that we are measuring the right thing.[x] From the latter viewpoint, precision is not just a useful thing to have, but it is a mark that we are closing in on the true regularities of nature. Metrologists are not merely been responding to physical theory in deciding to redefine the kilogram by fixing the value of the Planck constant, for metrology is concerned with revealing nature as she is, determining by scientific experiment what is the most uniform and immutable in nature. Although the most obvious sign of that a physical determination is accurate comes from coherence between two different methods of measuring it, it is also the case that precision indicates that a measure is—as metrologists tend to say—"more fundamental" than another (Mills *et al.*, 2011, 3). One way of acknowledging the dual aspect of the motivation behind the current SI reform is to accept that metrological progress is best captured by what is both a mild and a strong realism. Towards this end I propose the thesis of *metrological realism*: nature does not endorse all scientific standards equally; some scientific standards have the potential to be realised with lower uncertainties because they more closely reflect the regularities of nature.

Analysis of the theoretical desires behind the reform of the SI thus reveals, I believe, that the development of scientific standards is more closely associated with physical theory than the conventionalist account allows. There is no doubt that, in choosing a certain number, there is a conventional element to fixing the value of the Planck constant to that number. But the conventional aspect of this unit-setting is so remarkably tame, no more than is to be found elsewhere in the physical sciences, that it won't do to write of the kilogram itself as a convention. There is more to be said about the nature of this theoretical aspect of metrology, however. In the next section, I go on to consider two metrological procedures associated with the IPK: the current BIPM procedure for cleaning and washing of kilogram prototypes and the proposed use of the Watt balance to calibrate a mass standard. So far, I have been challenging the conventionalist distinction between accuracy and precision; I now look to take the challenge to the distinction between experiment and calibration and between a pointer and its *mise en pratique.*

4. The Watt balance and the metrological unification of the physical sciences

In 1975, Kibble (1976) proposed that a beam balance could be used to compare a gravitational power with an electrical one; the principle of the resulting Watt balance is illustrated in Figure 3. A mass m is suspended from one side of the balance; an electric coil of length L is suspended from the other in radial magnetic field of flux density B. To begin with, the balance is brought to equilibrium by passing a current I through the coil. The beam balances when the electromagnetic force BIL exerted upon one side of the beam matches the gravitational force mg upon the other:

$$mg = BIL$$

In the second phase of the procedure, the coil is moved downwards with velocity v, which induces an electric potential U, where $U = BLv$. Thus, the mass suspended from the balance can be expressed as a voltage and a current:

$$mgv = IU$$

Or, replacing the current I with U/R (by Ohm's law), where R is the resistance of the coil:

$$mgv = U^2/R \qquad [1]$$

This enables a mass measurement to be taken by reading an electromagnetic force. If U, R, g and v are measured in SI units, the resulting mass will be given in kilograms. A thorough description of the procedure is given by Kibble and Robinson (2003).

At first, the Watt balance was proposed as an improvement to the ampere balance and it did not have the promise of an atomic standard. It was the confirmation of the quantum Hall effect in the 1980s that brought this about. The Hall effect is the presence of a transverse voltage in an electrical conductor. Quantised, a resistance (the transverse resistance of the Hall effect) is expressed in terms of the Planck constant h, the elementary charge e and an integer p:

$$R = h/pe^2 \quad [2]$$

This allows for the resistance R of [1] to be expressed using microscopic quantities. The same can be done for the electric potential U of [1] by applying the theory of the Josephson effect, known since the 1960s. An electric potential exists at the junction of two superconductors separated by a thin layer of non-superconducting material and can be expressed in terms of the Planck constant, a frequency f (the phase difference between the metals), an integer q and the elementary charge e:

$$U = qfh / 2e \quad [3]$$

Thus a link can be made between a macroscopic mass and the Planck constant (inserting [2] and [3] into [1]):

$$m = q^2 f^2 he/4gv$$

As a result of advances in quantum electromagnetism, the Watt balance now provided a new experimental way to determine the Planck constant, by suspending a known mass upon it. But it also offered the possibility that, if the numerical value of h were fixed, it could be used to assign mass measurements.

Curiously, then, an experiment that was once used to determine the Planck constant will, after the current reform of the SI, act instead as a procedure to calibrate masses against the new definition of the kilogram. Thus, the choice of a metrology determines whether a laboratory procedure (the operation of the Watt Balance) is to be considered an experiment (to determine a physical constant) or a calibration (to determine the mass of an intermediate standard). This somewhat challenges the assumption that it is scientific experiments which provide science with its empirical content, in contrast to calibrations, which are mere translations of that content into a more convenient, universal language. A calibration is, after all, usually understood as a way of translating the indications given from laboratory equipment into a language understandable to those outside of that particular laboratory. As the BIPM define it (2012, 28):

> "an operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication".

Without denying the physical necessity of talking in kilograms instead of the degrees of the pointer of the red balance in the Mott Building, the conventionalist account of scientific standards marks the moment when the position of the red-balance pointer was recorded as bringing the empirical part of the experiment to an end with respect to the mass measurement. Each experiment in the physical sciences is supported by a collection of calibrations, which are each supported by further calibrations, ultimately leading to one or more of the *mise en pratiques* of the SI standards, in a way that can be depicted by a spider-web diagram. From this viewpoint, it comes as a surprise that a reform of the SI units does not merely consist of supporting an experiment with alternative calibration chains, but redefines what counts as a calibration in the first place. It remains open, however, how deeply to interpret this observation and how far the conventionalist account of scientific standards must be altered in order to accommodate it, as is attested by the history of twentieth century analytic philosophy, for this is, of course, a striking example of an old issue.

Logical empiricists saw the crash between experiment and calibration most clearly from the opposite side of the road. They appreciated that scientific knowledge previously understood to be purely empirical included a certain element of convention. Most famously, they acknowledged that, in order to formulate special relativity, Einstein had assumed that when light makes a return trip, it travels at the same speed on both legs of the journey. Reichenbach, amongst others, understood this assumption to be a convention, what he called a "co-ordinative definition" (1928). He considered such definitions to be logically necessary in order to allow the empirical work of science to get underway. Taking this route, it is possible to interpret the new use of the Watt balance as altering only the linguistic framework in which science is conducted (Carnap, 1950). The statement, "The Planck constant is $6.626,069,57 \times 10^{34}$ joule second," is empirical in one metrology and definitional in another. In accepting this, the logical empiricists recognised a single, one-way reliance between the conventional and the empirical.

It would, of course, be surprising, having rejected logical empiricism more generally in philosophy of science, to accept it as a success for metrology. And with hindsight, the first acknowledgement regarding the interplay between the conventional and the empirical is only the beginning of a slippery slope. I leave it open, however, how far a Quinean turn needs to be taken at this point. My sense is only that, after the toppling of logical empiricism, later philosophy of science has not been articulated with metrology in mind. It remains to test holistic theses in a metrological setting. Curiously, in this specific context, the associations between experiments marked by calibrations can be distinguished and separated from the larger web of knowledge and, furthermore, translation occurs here *within* this web. It raises the question, then: what exactly is a calibration? The case for a Quinean metrology is yet to be made, for it is apparent that the analytic-synthetic distinction still remains deeply imbedded in thinking about measurement. Quine's philosophy is yet to be applied to the science of measurement and brought to bear upon the specifics of metrological practice. One suggestion in that regard is to indeed consider a more contained holism than Quine's, which can therefore recognise the diversity of the sciences and the compartmentalisation of knowledge more generally. In this regard, I propose that it may sometimes be useful to consider the physical sciences as the collection of all determinations of physical quantities, regardless of whether they are, at any particular time, labelled as calibrations or experiments, and which are associated (in a theory-laden way), because some determinations make use of others. The metrological unification of the physical sciences gives metrology a similar epistemic status to other sciences, viewing the totality of determinations as providing the physical sciences with its empirical content. My sense is that the unification of the physical sciences is worth emphasising today because of the philosophical community's growing understanding of the diversity of the sciences in general. I am not arguing against the seemingly obvious truth that the physical sciences are united by a number of natural laws that connects variables representing different kinds of quantity, but merely emphasising that metrology is not merely an effort to express those equations with units, but takes its part in determining which connections are to be regarded as definitional, and which as holding empirical content. My thesis is not that metrology is everything in the physical sciences, but that it is more than current philosophical thinking allows for.

Although I do not suspect that a holism about language in general will prove very useful to the philosophy of measurement, it remains the case that if we wish to interpret the status of the Watt balance experiment from a conventionalist viewpoint we find ourselves between a rock and a hard place. The conventionalism of the logical empiricists diverges from the conventionalist account of scientific standards in two important ways. Although Reichenbach would agree that the choice of a scientific standard is a convention, he has reversed the order of communications between the metrologist and the physical scientist. On the logical empiricist account, co-ordinative definitions are necessarily prior to physical research; on the conventionalist account of scientific standards, as I have presented it, metrological pointers are handy aids for experimenters, but are necessarily informed by physical theory. The scientific-standard conventionalist more closely describes the case of the kilogram reform here, given its dependence upon quantum electrodynamics. It is widely recognised, of course, that the logical empiricists failed to account for the extent with which observation is laden with theory. The trouble with Reichenbach's co-ordinative definitions is that they are improved by empirical research. Secondly, although it is not explicit in Reichenbach's discussion of co-ordinative definitions, it is reasonable to interpret his view as holding little distinction between a scientific standard's pointer and its method of measurement. In the next section (§ 5), we shall see that it is the scientific-standard conventionalist who is at a disadvantage here.

Both theories do best where they make small steps to allow for the interplay between the empirical and the conventional, the theoretical and the empirical, the work of the metrologist and that of the physical scientist. There is room on the philosophical landscape to draw up a metrological account of scientific standards that more keenly appreciates the depth of this interplay and yet which perhaps does not discard these distinctions altogether. I take it that the most important steps in this direction have been made by Hasok Chang in his detailed historical accounts of progress in thermometry (2004). Each of Chang's case studies starkly demonstrates the empiricist's difficulty of making metrological progress because, on the face of it, there is nothing better to test our best scientific standards against but those standards themselves. Steering between conventionalism regarding scientific standards and operationalism, Chang outlines his theory of epistemic iteration, a version of coherentism, to account for how metrologists do, in practice, overcome this apparent circularity. Unlike the interval scales of early thermometry, the mass scale is a ratio scale, requiring (at least theoretically) just one fixed point. The underlying difficulty remains, however, that we wish to experimentally determine whether our chosen point is truly stable. Thus, the case study presented here is of the same ilk: how can we experimentally determine the most reliable measure of mass, without already having the most reliable measure of mass at hand? In today's reform of the kilogram, the solution has been partly provided by theoretical physics. It is partly on account of Chang's work that I reject the naively realist viewpoint that supposes metrology can rest on a foundation given by the constants of nature. It remains, after all, for metrologists to perform further precise measurements, testing and redefining scientific standards in the future. I have argued, however, that the coherentist has to accept a certain amount of realism in order to account for the intricacies of metrological history. I go on now to argue that, similarly, there is no need to reject all the tenets of operationalism, but merely to tame them.

5. The cleaning and washing of the IPK and the case for metrological operationalism

In preparation for each of the periodic verifications, the kilogram prototypes, including the IPK itself, were cleaned by a procedure documented by the BIPM (Thiesen 1889; CIPM 1946; Girard, 1990). For the third periodic verification, each prototype was first rubbed with a chamois leather cloth, which had been soaked three times, each for 48 hours, in a mixture of equal parts ethanol and ether, before the solvent was wrung out. A fairly hard pressure, estimated to be in the region of 10^4 Pa, was applied during the rubbing. Next, the prototype was steam washed to remove all traces of the solvent. For this, a jet of steam was sprayed directly at the surface of the prototype from an orifice of diameter 2 mm and at a distance of approximately 5 mm away. Any remaining water droplets were removed using an edge of filter paper.

Although a similar procedure had been performed before the first and second verifications, the third verification included an investigation, also conducted by Girard, into the effects of cleaning with solvent and steam washing upon the prototypes. Girard's results are shown in Figure 4, revealing that, in the months after being cleaned and washed, the mass of the platinum-iridium prototypes increased by 1 μg per month. As a result of this work, it was appreciated that a clarification to the 1889 definition of the kilogram was required. In 1989, the CIPM confirmed that the original definition of the kilogram referred to the IPK just after washing and cleaning by the official BIPM procedure; any comparison made to the IPK would therefore have to include an extrapolation to this mass. When accepting this proposal, the committee made it clear that this was not be interpreted as an alteration to the definition of the kilogram standard itself: "After considerable discussion, the Comité adopted this interpretation for the purposes of the third verification but made it clear that this did not in any way constitute a redefinition of the kilogram" (1989, 104). It did, however, enter the BIPM's SI brochure, included in the instructions for realising a kilogram (2006, Appendix 2).

The clarification highlights the current impossibility of knowingly obtaining a complete pointer (or *mise en pratique*) for a scientific standard. The metrological community is aware that a pointer is not entirely given by a short description and is willing to assume that an additional interpretation is required. In this case, a clarification that could have been added to the definition of the kilogram was instead included in its *mise en pratique.* The history of metrology is littered with occasions when a definition was instead altered in order to clarify its determination.[xi] The logical empiricists were right, then, that the distinction between a pointer and its *mise en pratiques* is not as clear as it is presented by the conventionalist account of scientific standards. Historically, the decision to choose a *mise en pratique* is bound up with the choice of a pointer and it is not obvious where the line between the two should be drawn, if it is to have epistemic significance. This was appreciated by generic conventionalists and thus I conclude, with Reichenbach amongst others, that metrology takes its place in necessarily setting the framework for physical theory. Yet, as I have argued in the last section, generic conventionalism does not follow through the consequences of this point far enough. It is not just that metrology gives physical theory a conventional nature, but also the reverse: in addition, it performs an empirical role for the physical sciences. Echoing the point of § 3, the selection of a pointer and its realisations is an investigation into the nature of reality.

The point can be considered from a semantic angle: is it the definition of a scientific standard or the procedure to realise the definition that gives meaning to that standard? The logical empiricists emphasised the importance of physical procedures in conferring meaning upon words. More generally, twentieth century philosophers have suggested that meaning is to be found in the way a word is used, or the way a statement is verified, or the actions that are performed when applying a word.[xii] Even for the most practically minded philosophers, however, this doctrine has not been brought to bear in a way that hinges the meaning of a scientific concept on the full and intricate network of metrological procedures by which we measure quantities of that concept. At most, a verificationist goes as far as finding the meaning of, "It has a mass of so many kilograms," in the single, immediate procedure that a scientist would perform to check that result. In retrospect, an alternative is available to the philosopher seeking meaning in actions: the entire chain of calibrations, performed previously (and in most cases by other people), leading all the way back to the IPK, is where lies the meaning of "It has a mass of so many kilograms," and thus the significance of "kilogram" and then, perhaps, even "mass".

It is perhaps Percy Bridgman's writings, especially in his later work, that are most suggestive of this theory of meaning. He writes that, "the meanings of one's terms are to be found by an analysis of the operations which one performs in applying the term in concrete situations *or* verifying the truth of statements *or* in finding the answers to questions" (1938, 114–131, emphasis added). Bridgman thus allowed for many different kinds of operations (including those he described as "mental" and "of paper and pencil") to confer meaning, although he never explicitly offered metrological procedures to fulfil this role. In the case of mass, for example, Bridgman assumed that the concept was best understood by contemplating procedures to measure force in the absence of a gravitational field—his suggestion was the deformation of elastic materials—and then extricating the concept of mass from this (1927, 102–108). The proposal reveals a theoretical bias in even the most pragmatic of physicists. Why turn to Newton's equivalence between mass and force and not instead take the operations that are actually used in scientific practice to measure mass, which in Bridgman's time, as well as now, ultimately required the IPK, sitting upon a balance tray?

Bridgman did not intend to advocate any theory of meaning. A Nobel-prize winning physicist himself, he promoted his "operational attitude" as a way for colleagues to think more clearly about the concepts they dealt with. The kilogram never appeared on his radar. After all, it was obviously a term that lacked any conceptual confusion. Nor have scientific units, their manifestly conventional definitions being clearly stated and firmly agreed upon by the scientific community, appeared to philosophers to hold much philosophical promise.[xiii] Even the most anti-realist of thinkers have assumed a hierarchy between "mass" and "kilogram" (and thus accepted, unwittingly or not, a little realism): it is only the concept of mass that is troubled, once we define and understand what that is, whatever that may be, a kilogram is nothing more than a particular amount of it. Thus they have accepted the hierarchy of the conventionalist account of scientific standards: a definition of a unit is relegated below that of the quantity it measures. Anti-realism is thus forced to take on a two-sided conventionalism that rather mimics the realist position it was intended to oppose. Although the verificationalist or operationalist looks to procedures, actions or operations to reveal what mass is, these are separated from those that reveal what a kilogram is, the former being required prior to and independently of the latter. When the scientist, in a particular context, turns only to a particular metrological procedure to take a mass measurement, the verificationist looks to that procedure to provide only the meaning of "kilogram", and then looks through the textbooks to discover what could possibly be used to provide a further meaning for "mass". On reflection, this is a peculiar task because the metrological work that generates the kilogram also gives science its processes for measuring mass. Furthermore, in practice, we do not need to be committed to a particular view about mass, but only to make assumptions about what kinds of thing retain their mass over time, in order to define a mass unit. In practice, procedures to make measurements—at least preliminary ones—come before a theoretical understanding of what is being measured.

The operational attitude of Bridgman—as well as the like-minded anti-realism of the generic conventionalist—thus faces the problem that, despite first appearances, it has not achieved what it set out to do in overcoming the crude distinction of naïve realism between the objective and the subjective. It upholds the realist's hierarchy between a quality and its standard, using different operations to determine the meaning of each. To retain an internal simplicity and coherence to the thesis of operationalism—to go the whole way—it is necessary to look to all the metrological procedures associated with a standard to give meaning, not just to that standard, but the quality it represents. This alternative, more thoroughgoing operationalism reverses the dependence between "kilogram" and "mass". It overcomes one of the main difficulties facing Bridgman's theory, by clarifying what counts as an 'operation' (which is now understood as a procedure used by the scientific community, ratified by metrologists, to make measurements). In embracing all metrological procedures as holding significance, thoroughgoing operationalism makes little distinction between calibrations and the precision measurements of experiment. What is more, by more vigorously shaking off conventionalism regarding scientific standards, it provides a stronger base from which to respond to the problem of accounting for metrological progress, considered so far with respect to the scientific-standard conventionalist. The problem presents itself even more starkly against the operationalism of Bridgman. Donald Gillies (1972, 6–7) raised the point most clearly: if we declare the meaning of scientific measures to be found in the bare metrological procedures for their realisation, why would we ever want to improve a method of measurement?

A logical empiricist, verificationalist, operationalist or even a more thoroughgoing operationalist—appears to be truly stuck. In my interpretation of the issue in § 3, however, I have argued that the naïve realism associated with the conventionalist account of scientific standards faces the same issue. The fix is not to add a dose of realism about scientific posits alongside a conventional view of scientific standards, attempting to explain metrological progress as the move towards a truer or more objective measure. Metrologists are not redefining the kilogram after having been presented with a more accurate account of what mass is. It was their decision to reject alternative definitions of the kilogram. The 'thick' motivation of stability—the fact that the better definition is revealed by improved metrological precision—is not accounted for by half-half vision of science (what it measures is real; how it measures it is convention). Gillies' objection does not only apply to operationalism; neither does a simple realistic picture of science accurately portray the progress of metrology. What is more, I believe that the reason Gillies' argument is so arresting is that it supposes that operationalism adheres to certain elements of the conventionalist account of scientific standards. The operationalist is assumed to be a perfect conventionalist in that the operations chosen to provide meaning must necessarily be chosen for their practical advantages only. When we realise that these operations are providing meaning precisely because the assumptions implicit in those operations—the conservation of mass of an iridio-platinum piece, for example—are true to reality, we can see that an operationalist reforms a scientific standard to improve the stability of measurement, a notion that is at once pragmatic and realist.

Thoroughgoing operationalism is far too cumbersome as it stands, however. I do not advocate it in its entirety but tame it as much as I can: *metrological operationalism* is the thesis that the meaning of a scientific concept is at least partly given by the collection of all methods by which it is realised, including the *mise en pratiques* for its scientific standard. This is the semantic version of the thesis that metrology contributes empirical content to science; I propose it as a possible contribution to a metrological account of scientific standards.

6. A metrological account of scientific standards

The discussion of this paper intended to motivate the larger question: what role do scientific standards play in the development of physical theory? I take a metrological account of scientific standards to be a response to this that takes into consideration the practice, both current and historical, of metrology. I have argued that this is necessary because the conventionalist account of scientific standards, most clearly associated with realist thinking but also identifiable in anti-realist thought and reaching further into our philosophical thought than is immediately apparent, fails to account for metrological progress. I have made tentative suggestions for a metrological account of scientific standards. It might reasonably include a version of metrological operationalism (but would in any case acknowledge that the meaning of a standard is not wholly contained in its official definition). It might reasonably support metrological realism regarding scientific standards (but would in any case acknowledge a deeper interplay between metrology and physical theory than conventionalist and realist thinking). It might reasonably view the physical sciences to be unified by metrology (but would in any case acknowledge that metrology sets the empirical limits of the physical sciences or even contributes to its empirical content).

The overlaying of the history of metrology upon that of twentieth century analytic philosophy that I have applied here is not merely a series of analogies (between the results of calibration and analytic statements, between the results of experiment and synthetic statements, between the procedure of translation and those of calibration). Metrology offers a testing ground for theories of meaning. The conventionalist account of scientific standards has stood in the way of undertaking real metrological examples to philosophy of science and philosophy more generally. Elements of this account resonate in philosophical thought, despite the fact that it is at odds with more general conventionalist thinking. Reichenbach was perfectly right to point out particular conventions are necessary to get measurement off the ground and ultimately sat disguised within the body of physical theory. It is not an observation that generalises well, however: the resulting conventionalism ultimately had the unfortunate side-effect of sweeping metrology beneath the carpet. The search for convention entering into theory and experiment should be complemented with a search for the theoretical and empirical resulting from measurement. (The point, after all, has always been that "theoretical" and "conventional" are not terms that stick fast in philosophy.) Here, I have attempted to work towards that goal, in a way that echoes Peter Galison's focus on an experiment's end.

Despite the lack of a deductive closure to experiment, Galison argued that rising experimental evidence eventually becomes sufficiently persuasive to draw an experiment to an end (1988). He argued that the complicated tangle of factors that brought experiments to this point had to be unravelled from a historical perspective, discouraging simplistic philosophical models. Mimicking the logician's crude closure of an experiment, the scientific-standard conventionalist assumes simple distinctions (between an experiment and its associated calibrations, between the empirical nature of physical theory and the conventional nature of metrology, between the accuracy of the idealist and the precision of the pragmatist) to mark the beginning of an experiment. It is only after we know what we are measuring and how we are measuring it, that experimental science may begin. I've been arguing here, then, that a more detailed, historical approach is also required in order to untangle the factors which bring experiments to their beginning.

Acknowledgements

I would like to thank two anonymous reviewers for their particularly insightful comments. I would also like to thank Jeremy Butterfield, Hasok Chang, Jim Grozier, Eran Tal, as well as all other members of the Philosophy and History of Science reading group at the University of Cambridge. I thank David Lloyd for the creation of Figure 3.

References

Avetissian, A. K. (2009), Planck's constant variation as a cosmological evolution test for the early universe, *Gravitation and Cosmology*, 15, 10–12

Ayer, A. J. (1936). *Language, Truth and Logic.* London: V. Gollancz, Ltd.

Bell, S. (2001). A Beginner's Guide to Uncertainty of Measurement, *Measurement Good Practice Guide,* No. 11, Issue 2, National Physical Laboratory.

Blevin, W. R. and Steiner, B. (1975). Redefinition of the Candela and the Lumen, *Metrologia* 11, 97–104, doi:10.1088/0026-1394/11/3/001.

BIPM (2006). *SI Brochure: The International System of Units (SI)*, 8th edition, updated in 2014, Bureau international des poids et measures.

BIPM (2008). Evaluation of measurement data—Guide to the expression of uncertainty in measurement, *Joint Committee for Guides in Metrology*, 100.

BIPM (2012). International vocabulary of metrology – Basic and general concepts and associated terms (VIM), *Joint Committee for Guides in Metrology*, 200.

BIPM (2013). *Draft Chapters 1, 2 and 3 of the 9th SI Brochure*, Bureau international des poids et measures.

Bridgman, P. W. (1927). *The Logic of Modern Physics.* New York: Macmillan

Bridgman, P. W. (1938). Operational Analysis, *Philosophy of Science* 5, 114–131.

Carnap, R. (1950). Empiricism, Semantics, and Ontology, *Revue Internationale de Philosophie*, 4, 20–40.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford Studies in the Philosophy of Science. New York: Oxford University Press.

Condamine, C. M. de La (1747). *Nouveau projet d'une mesure invariable propre à servir de mesure commune à toutes les nations*, Académie royale des sciences.

CIPM (1946). *Procès-verbaux*, Series 2, Vol. 20, Comité international des poids et mesures.

CIPM (1989). *Procès-verbaux de la 78e session*, Vol. 59, 1989, Comité international des poids et mesures.

CIPM (1993). *Procès-verbaux de la 82e session*, Vol. 61, 1993, Comité international des poids et mesures.

CIPM (2008). *Procès-verbaux de la 97e session*, 2008. Comité international des poids et mesures.

CGPM (1890). *Comptes rendus des séances de la première conférence générale des poids et mesures, réunie à Paris en 1889*, Paris: Gauthier-Villars.

CGPM (1928). *Comptes rendus des séances de la septième conférence générale des poids et mesures, réunie à Paris en 1927*, Paris: Gauthier-Villars.

CGPM (1949). *Comptes rendus des séances de la neuvième conférence générale des poids et mesures, réunie à Paris en 1948*, Paris: Gauthier-Villars.

CGPM (1975). *Comptes rendus des séances de la quinzième conférence générale des poids et mesures, Centenaire de la Convention du Mètre et du Bureau International des Poids* *et Mesures, Paris 27 Mai–2 Juin 1975*, Paris: Bureau International des Poids et Mesures.

CGPM (2010). *Comptes rendus de la 23e réunion de la Conférence générale des poids et measures (novembre 2007),* Paris: Bureau International des Poids et Mesures.

Davies, R. S. (2005). "Possible new definitions of the kilogram", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363, 2249–2264. doi:10.1098/rsta.2005.1637

Dirac, P. A. M. (1937). The Cosmological Constants, *Nature* 139, 323, doi:10.1038/139323a0.

Dummett, M. (1978). *Truth and Other Enigmas*. London: Duckworth.

Galison, P. (1987). *How experiments end.* Chicago, Illinois: University of Chicago Press.

Gill, D. (1907). Address of the President of the British Association for the Advancement of Science, *Science*, New Series, Vol. 26, No. 659, pp. 193–212.

Gillies, D. (1972). Operationalism, *Synthese* 25, pp. 1–24.

Girard, G. (1990). *The procedure for cleaning and washing platinum-iridium kilogram prototypes used at the Bureau International des Poids et Mesures*, Bureau International des Poids et Mesures.

Hill, T. P.; Miller, J.; and Censullo, A. C. (2011). Towards a better definition of the kilogram, *Metrologia* 48, 83–86 doi:10.1088/0026-1394/48/3/002.

Huygens, C. (1673). *Horologium oscillatorium: sive, de motu pendulorum ad horologia aptato demostrationes geometricae*, Paris: F Muguet.

Lavoisier, A. L. (1893). Eclaircissements historiques sur les mesures des anciens, in M. Dumas et E. Grimaux (eds.), *Oeuvres de Lavoisier*, 1864–1893, Vol. 6, pp. 698–705. Paris: Imprimerie Nationale.

Kelley, M. H. (2001). The Electronic Kilogram, *Microwave Symposium Digest, IEEE MTT-S International*, pp. 857 – 860.

Kershaw, M. (2012). The 'nec plus ultra' of precision measurement: Geodesy and the forgotten purpose of the Metre Convention, *Studies in History and Philosophy of Science Part A* 43 (4) pp. 563–576.

Kibble, B. P. (1976). A measurement of the gyromagnetic ratio of the proton by strong field method, *Atomic Masses and Fundamental Constants,* 5, J. H. Sanders and A. H. Wapstra, Eds. New York: Plenum, pp. 545–551.

Kibble, B. P. and Robinson, I. (2003). Replacing the Kilogram, *Measurement Science and Technology*, 14, 1243–1248.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Mavidal, J. and Laurent, E. (1881). *Archives parlementaires de 1789 à 1860; recueil complet des débats législatifs & politiques des chambres française*. Series I, Vol. 12, Paris: Dupont.

Maxwell, J. C. (1870). Address to the Mathematical and Physical Sections of the British Association, *Report of the Fortieth Meeting of the British Association for the Advancement of Science*, 40, Notices and Abstracts, pp. 1–8.

Mills, I. M., Mohr, P. J., Quinn, T. J., Taylor, B. N. and Williams, E. R. (2011). Adapting the International System of Units to the twenty-first century, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369, 1953. pp. 3907–3924. doi:10.1098/rsta.2011.0180

Mulligan, J, (1998). Who were Pérot and Fabry?, *American Journal of Physics*, 66, 9.

Nagel, E. (1979). *The structure of science.* Indianapolis: Hackett.

Picard, J. F. (1671). *Mesure de la terre*, Paris.

Quine, W. V. O. (1960). *Word and Object*, Cambridge, Mass.: M.I.T. Press.

Quinn, T. (2011). *From Artefacts to Atoms: The BIPM and the Search for Ultimate Measurement Standards*, Oxford University Press.

Reichenbach, H. (1928). *Philosophie der Raum-Zeit-Lehre*. English translation: Maria Reichenbach, 1957, *The Philosophy of Space and Time*. Dover.

Riordan, S. (2015). The Objectivity of Scientific Measures, *Studies in History and Philosophy of Science*, 50, pp. 38–47.

Schaffer, S. (1995). Accurate Measurement is an English Science, *Values of Precision*, ed. M. Norton Wise, pp 135–172. Princeton University Press.

Stock, M. (2011). The watt balance: determination of the Planck constant and redefinition of the kilogram, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. doi: 10.1098/rsta.2011.0184

Talleyrand, C. M. (1790), *Proposition faite a l'assemblee nationale sur les poids et mesures par M l'evê ue d"Autun.* Paris: L'Imperimerie national.

Thiesen, M., (1898), *Travaux et mémoires du Bureau international des poids et measures*, 9.

Whitehurst, J. (1787). *An Attempt Towards Obtaining Invariable Measures of Length, Capacity, and Weight, From the Mensuration of Time, Independent of the Mechanical Operations Requisite to Ascertain the Center of Oscillation, or the True Length of Pendulums*. London: W. Bent.

Wilkins, J. (1688). "Of Measure", from *An Essay Toward a Real Character and a Philosophical Language*, London: S. Gellibrand.

Wittgenstein, L. (1953). *Philosophical Investigations*, G. E. M. Anscombe and R. Rhees (eds.), G. E. M. Anscombe (trans.), Oxford: Blackwell.

Captions

Tab. 1. Currently accepted values of the physical constants that will be fixed by the latest reform of the SI (BIPM, 2013). © Bureau International des Poids et Mesures. Reproduced by permission of the BIPM. All rights reserved. *(Located in § 1.)*

Tab. 2. Results of the third periodic verification of the national prototypes (CIPM, 1993, G43). © Bureau International des Poids et Mesures. Reproduced by permission of the BIPM. All rights reserved. *(Located in § 3.)*

Fig. 1. Change in mass of the national prototypes No. 2 to 20 (those made prior to the first verification of 1889), as well as official copies No. 8(41) and No. 32, with respect to the IPK (CIPM, 1993, G45). © Bureau International des Poids et Mesures. Reproduced by permission of the BIPM. All rights reserved. *(Located in § 3.)*

Fig. 2. Change in mass of the national prototypes No. 44 to 55 (those made after the first verification of 1889 but before the second of 1946), as well as official copies No. 8(41) and No. 32, with respect to the IPK (CIPM, 1993, G45). © Bureau International des Poids et Mesures. Reproduced by permission of the BIPM. All rights reserved. *(Located in § 3.)*

Fig. 3. The principle of the Watt balance: (a) when in equilibrium; (b) when in motion. *(Located in § 4.)*

Fig. 4. The change in mass of the kilogram prototypes on cleaning and washing, plotted against the years elapsed since their last cleaning and washing. Crosses identify the IPK and its six official copies. Open circles indicate prototypes with poor surface condition (Girard, 1990; CIPM, 1989, 130). © Bureau International des Poids et Mesures. Reproduced by permission of the BIPM. All rights reserved. *(Located in § 5.)*

Tab. 1

Physical constant	**2010 CODATA value**
the unperturbed ground state hyperfine splitting frequency of the caesium 133 atom $(_{133}Cs)_{hfs}$	9,192,631,770 hertz
the speed of light in vacuum *c*	299,792,458 metre per second
the Planck constant *h*	$6.626{,}069{,}57 \times 10^{34}$ joule second
the elementary charge *e*	$1.602{,}176{,}565 \times 10^{19}$ coulomb
the Boltzmann constant *k*	$1.380{,}648{,}8 \times 10^{23}$ joule per Kelvin
the Avogadro constant N_A	$6.022{,}141{,}29 \times 10^{23}$ reciprocal mole
the luminous efficacy K_{cd} of monochromatic radiation of frequency 540×10^{12} hertz	683 lumen per watt

Tab. 2

International prototype	𝕶		1 kg
Official copies	K1		1 kg + 0,135 mg
	No. 7		1 kg – 0,481 mg
	No. 8(41)		1 kg + 0,321 mg
	No. 32		1 kg + 0,139 mg
	No. 43		1 kg + 0,330 mg
	No. 47		1 kg + 0,403 mg
BIPM prototypes	No. 25		1 kg + 0,158 mg
	No. 9		1 kg + 0,312 mg
	No. 31		1 kg + 0,131 mg
National and other prototypes	No. 2	Romania	1 kg – 1,127 mg
	No. 5	Italy	1 kg + 0,064 mg
	No. 6	Japan	1 kg + 0,176 mg
	No. 12	Russian Federation	1 kg + 0,100 mg
	No. 16	Hungary	1 kg + 0,012 mg
	No. 18	United Kingdom	1 kg + 0,053 mg
	No. 20	United States of America	1 kg – 0,021 mg
	No. 21	Mexico	1 kg + 0,068 mg
	No. 23	Finland	1 kg + 0,193 mg
	No. 24	Spain	1 kg – 0,146 mg
	No. 34	Académie des Sciences de Paris	1 kg – 0,051 mg
	No. 35	France	1 kg + 0,189 mg
	No. 36	Norway	1 kg + 0,206 mg
	No. 37	Belgium	1 kg + 0,258 mg
	No. 38	Switzerland	1 kg + 0,242 mg
	No. 39	Rep. of Korea	1 kg – 0,783 mg
	No. 40	Sweden	1 kg – 0,035 mg
	No. 44	Australia	1 kg + 0,287 mg
	No. 46	Indonesia	1 kg + 0,321 mg
	No. 48	Denmark	1 kg + 0,112 mg
	No. 49	Austria	1 kg – 0,271 mg
	No. 50	Canada	1 kg – 0,111 mg
	No. 51	Poland	1 kg + 0,227 mg
	No. 53	Netherlands	1 kg + 0,121 mg
	No. 54	Turkey	1 kg + 0,203 mg
	No. 55	Fed. Rep. of Germany	1 kg + 0,252 mg
	No. 56	South Africa	1 kg + 0,240 mg
	No. 57	India	1 kg – 0,036 mg
	No. 58	Egypt	1 kg – 0,120 mg
	No. 60	People's Rep. of China	1 kg + 0,295 mg
	No. 65	Slovak Rep.	1 kg + 0,208 mg
	No. 66	Brazil	1 kg + 0,135 mg
	No. 68	Dem. People's Rep. of Korea	1 kg + 0,365 mg
	No. 69	Portugal	1 kg + 0,207 mg
	No. 70	Fed. Rep. of Germany	1 kg – 0,236 mg
	No. 71	Israel	1 kg + 0,372 mg
	No. 72	Rep. of Korea	1 kg + 0,446 mg
	No. 74	Canada	1 kg + 0,446 mg
	No. 75	Hong Kong	1 kg + 0,132 mg
	No. 3	Spain	1 kg + 0,077 mg
	No. 62	Italy (IMGC)	1 kg – 0,907 mg
	No. 64	People's Rep. of China	1 kg + 0,251 mg

Fig. 1

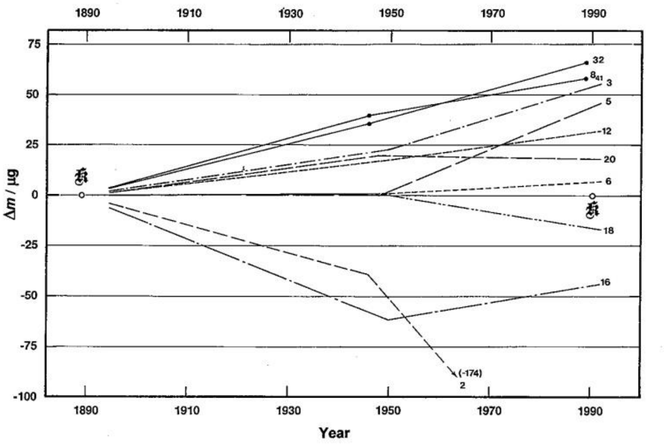


Fig. 2

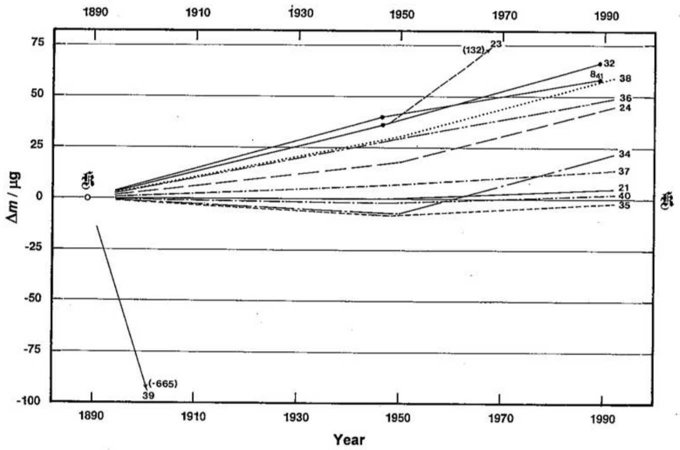


Fig. 3

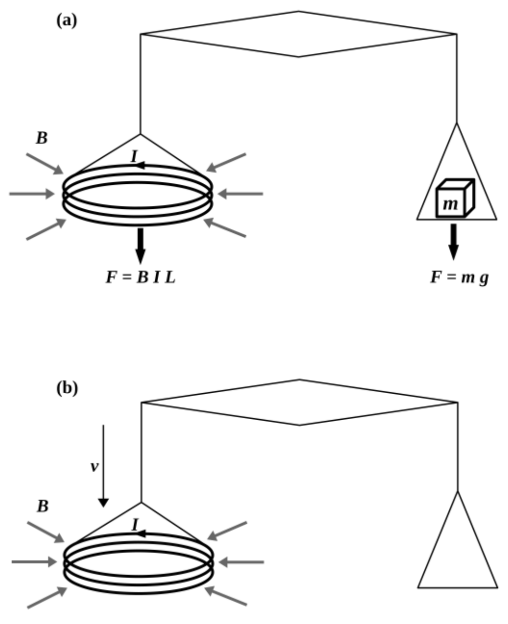
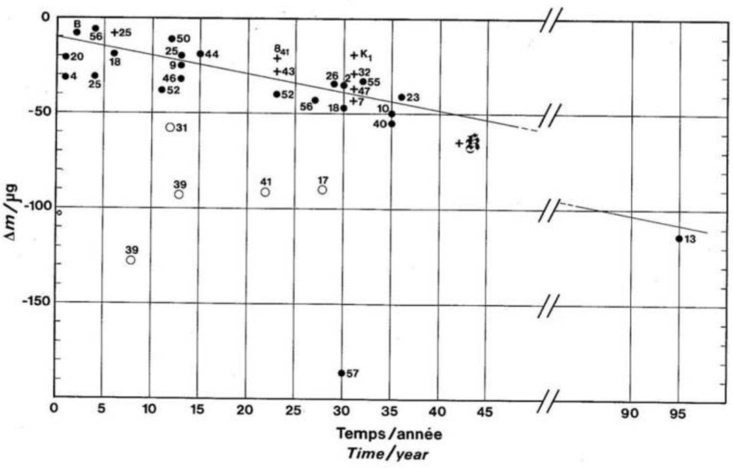


Fig. 4



[i] The motivations for the creation of the original kilogram are explored by Riordan (2015).
[ii] The motivations for the creation of the original metre are explored by Kershaw (2012).
[iii] The state of measurement in England at this time is surveyed by Schaffer (1995).
[iv] No 25 and No. 31 have since been renumbered.
[v] This hypothesis has been articulated by Terry Quinn, director of the BIPM between 1988 and 2003 (2011, 365).
[vi] Natural philosophers who supported using more fundamental definitions of scientific standards include Picard (1671), Huygens (1673), Wilkins (1688), de La Condamine (1747), Whitehurst (1787) and Lavoisier (1893). The metric project was initiated by a call for natural scientific standards, recorded by the editors of the parliamentary archives, Mavidal and Laurent (1881, 104–108). The resolutions of the CGPM show the desire for “a natural and indestructible standard” (1961, 85), “a natural base” (1949, 44), and to define scientific standards “in terms of the invariants of nature” (2010, 434). The desire is also reflected in articles of modern metrology, including Blevin and Steiner (1975), Kibble and Robinson (2003) and Mills *et al.* (2011).
[vii] A survey of the changing meaning of a fundamental standard is given by Riordan (2015).
[viii] Joseph Mulligan gives a brief account of the lives and work of Pérot and Fabry (1998).
[ix] De Jacobi, member of the Imperial Academy of Science, Saint Petersburg, brought up the possibility of using a collection of artefacts to define the kilogram at a meeting of the International Metre Commission in 1872 (Quinn, 2001, 52).
[x] I am appropriating the notion of a thick concept used in the ethics of science, in which a factual element as well as one of value can be found, a discussion that goes back to Ernest Nagel (1979, 485–502).
[xi] Examples include the 1927 amendment to the metre described in § 1.
[xii] There are many philosophical works that could be used to support my claim here. I’m thinking in particular of those of A. J. Ayer (1936), P. W. Bridgman (1938), Ludwig Wittgenstein (1953), W. V. O. Quine (1960) and Michael Dummett (1978).
[xiii] A notable exception is Wittgenstein’s insistence that it cannot be said of a metre stick that defines the metre that, “it is one metre long” (1953, §50) and Kripke’s ensuing suggestion that such statements are *a priori* contingent truths (1980).